DOCUMENT RESUME

ED 218 310 TM 820 350

AUTHOR Huntley, Renee M.; Plake, Barbara S.
TITLE An Investigation of Student Sensitivity to Cues in a Grammatically Consistent Stem and Set of Alternatives.
PUB DATE Apr 81
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cues; Grammatical Acceptability; Higher Education; *Responses; Response Style (Tests); *Student Reaction; *Test Format; Test Items; *Test Wiseness

ABSTRACT

This investigation of the effect of making a set of alternatives conform grammatically to a test item stem showed there is subject sensitivity to such cues. Content-free versions of American College Testing Assessment Experimental Social Science items representing singular-plural and vowel-consonant agreement without inappropriate grammatical situations were examined using freshman university students. Rephrasing items into neutral, but still precise and grammatically correct, language was found to eliminate sensitivity. Previous cueing research is presented in defining the present study. Two analyses of cue sensitivity are described. An appendix presents examples of the item types used in the study. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED218310

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

An Investigation of Student Sensitivity to Cues in a Grammatically Consistent Stem and Set of Alternatives

Renee M. Huntley
American College Testing Program
Barbara S. Plake
University of Nebraska-Lincoln

Paper presented at the annual meeting of the
American Educational Research Association,
Los Angeles, 1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. M. Huntley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 350

Abstract

This study investigates the effect of adhering to the universally endorsed item-writing practice of making a set of alternatives conform grammatically to a stem. Formats were designed to investigate whether the practice of providing relevant grammatical cues in itself created cue sensitivity in content-free items. Items are content-free versions of ACT Assessment Experimental Social Science items. The items represented grammatical examples of two types: singular-plural and vowel-consonant agreement. In no instance were grammatically inappropriate situations created. Results suggest subject sensitivity to cues. Rephrasing into neutral, but still precise and grammatically correct, language can eliminate this sensitivity.

An Investigation of Student Sensitivity to Cues in a Grammatically Consistent Stem and Set of Alternatives

Introduction

Test developers are warned to avoid specific flaws in writing test items because flaws influence test scores (Ebel, 1972). Among the elements of test-wiseness identified by Millman, Bishop, and Ebel (1965) to be dependent upon the test constructor were cue-using strategies of not making the alternatives "grammatically consistent with the stem." In his review of test-wiseness in the cognitive domain, Sarnacki (1979) points out that one effect of this item-writing fault is higher test scores. Dunn and Goldstein (1959), McMorris, Snyder, Brown, and Pruzek (1972) found that flaws in grammar produced easier items, while Board and Whitney (1972) failed to find significant effects. One strategy adopted by test developers to ensure grammatical consistency between the stem and the set of alternatives has been that of "generalizing" the stem: plural verbs or vowel-matching articles are added, usually parenthetically, to provide for grammatical agreement with all options. If test developers adopt this strategy selectively (that is, because the correct answer, in fact, is the one that differs grammatically from the alternatives in the set), an additional source of item cueing may be created that will affect item performance. Operationally, the very presence of a generalized stem in an item may establish a response set in the examinees that alerts them to select the alternative consistent

with the grammatical adjustment (a subtle cue) present in the stem. Sensitivity to legitimate or appropriate language cues in items may be a function of the personal characteristics of examinees--e.g., race, sex, grade level, anxiety level (Slakler, Koehler, & Hampton, 1970; Millman & Setijadi, 1966).

Huntley and Plake (1980) investigated the potential cueing that is created in singular-plural and vowel-consonant items when grammatical agreement between stem and alternatives is ensured by adding (usually parenthetically) to the stem plural verbs or vowel-matching articles. Some evidence by sex of differential sensitivity to these subtle cues was found. The items used in the study, however, had correct alternatives. The alternative-selection behavior of the students, therefore, could have been more a function of knowledge rather than of cue sensitivity, thus producing a source of confounding in the results.

Method

Two item types are considered in this study: vowel-consonant and singular-plural. (Examples of these items can be found in Appendix 1, page 18.) Sixteen items were selected: 9 measured plural-singular agreement, and 7 measured vowel-consonant agreement. The items were content-free versions of an experimental social science test. They had been administered in their content-true form as part of the ACT Assessment tryout. From this data base, pilot difficulty and discrimination indices were established.

For this study, two versions of each item were developed and labeled A and B. Version A items contained a generalizing grammatical cue

that augmented the stem with plural verbs or vowel-matching articles. In Version B, the set of alternatives for each item is the same as in its Version A counterpart; however, the stem of each item is rephrased to remove the generalizing cues. This rephrasing presents the stem in neutral, but still precise and grammatically correct, language.

Items that measure plural-singular agreement were unique in that they were based on a "combination" format. The combination item format represents a logical extension to the typical specific-answer multiple-choice item. The alternatives group the options (usually identified as I, II, III) in specific combinations. The student confronts a more complex task since the most comprehensively correct combination of options must be selected. No research on the combination format was found, although items of this nature appear in many standardized tests. Inclusive or exclusive options ("all. . ." or "none of the above") increase item difficulty (Dudycha & Carpenter, 1973; Hughes & Thumber, 1956), but whether this generalizes to combination items is yet to be studied.

The items were arranged into two experimental units in such a fashion that each student had one of the two versions of each item and the two item types were represented in each unit.

Procedure

The two experimental units were administered to a total of 231 freshman students who were enrolled in a Foundations of Modern Education class at a large midwestern university in the fall of 1980. The materials were administered in intact classrooms. For each class, the

students were informed that the study investigated whether the phrasing of multiple-choice questions would affect their performance. They were told that the subject area was social studies, that since they might not have had a recent course in social studies they might be unsure of correct answers for some of the questions. Although the students were encouraged to answer every question, even if they were unsure of the correct answer, they were advised not to guess randomly. Instead, they were asked to use the kinds of strategies they usually employ in a testing situation when answering multiple-choice questions for which they did not know the right answer. A description of the subjects across the two experimental units, broken down by sex, can be found in Table 1. A mean item score was calculated by scoring A = 1, B = 2, C = 3, D = 4 and then averaging the scores for males and females within each unit. The items were then sorted into the appropriate research form: A, which contained the generalizing cue in the stem and B, which contained the stem expressed in neutral language.

Analysis

Two distinct analyses were selected for analyzing different aspects of cue sensitivity. A general description of each analysis is provided first, after which the results are reported. (Plural-singular and vowel-consonant item types were analyzed separately; results for each are reported within each analysis performed.)

The procedure used in developing the items for the two research forms consisted of systematically rewriting the same item to match the

specifics of each form. The same item content, therefore, was varied across the forms. For plural-singular item types, 9 items were written two times; 7 vowel-consonant item types were also systematically varied in the same manner. Since the procedure systematically varied the presentation of the same item in two ways, an analysis of trends in performance across the forms was chosen. A two factor ANOVA design, with form considered a within factor and sex perceived as a between factor, was selected to analyze trends in the mean item scores across the two alternative versions.

Whether the presence or absence of the generalizing cue influences students' item-taking behavior was investigated by another series of analyses. The pattern of responses across the four alternatives was compared to a random pattern by means of a X^2 Goodness of Fit for each item type (plural-singular and vowel-consonant). This analysis was performed separately for plural-singular and vowel-consonant items.

The potency of the presence of the cues in the forms to direct student responses was investigated by considering the popularity of consistent alternatives in both situations (with the cue present in the stem and with the cue absent). The proportion of students selecting the consistent alternatives was compared by calculating the average shift to the consistent alternative(s) for each of the item types considered.

Results

Form Analysis

Item score means and standard deviations, by form and sex, for plural-singular and vowel-consonant item types are shown in Table 2. Results of the mixed model ANOVA (form, within; sex, between) are summarized in Table 3. No significant effects were found.

Cue Sensitivity

The presence of the subtle cue in the stem in Form A, if followed systematically by the examinees, would lead them to a set of alternatives consistent with the cue. If the students were sensitive to the cue, then the proportion of students selecting consistent alternatives would be reflected in the pattern of responses across the alternatives. This pattern of responses was compared to a random pattern for both types of items. The results for plural-singular item types for the total group and for males and females separately are shown in Table 4. They do not suggest a systematic pattern atypical from random.

For the vowel-consonant item types, one alternative is consistent with the presence of a generalizing cue, the alternative that begins with a vowel. The location of that alternative in the set of alternatives (A to D) varied randomly within the set of items. For the analysis of the distribution of alternative choices, the location of the consistent alternative was arbitrarily changed to location D. For each item, the distribution of percent of alternative choices was determined and averaged for consistent and other alternatives. These distributions were compared

to a random pattern using a X^2 Goodness of Fit test. The results are shown in Table 5. The X^2 values for males, females, and total group surpass the $.95^{X^2_3}$ critical value of 7.82. Therefore, the pattern of selection of alternatives does support a systematic influence of the generalizing cue.

Plural-singular type items in Form A contain one alternative that is consistent with the generalizing cue present in the stem. The proportion of students choosing this alternative was determined for each of the nine items. The same record was made of the same alternative for those items in Form B. The difference by item was calculated for the proportion choosing alternatives consistent from Form A to Form B. The average change in this proportion was 0.64, a negligible shift.

For the seven vowel-consonant type items, a similar computation was made that noted and then subtracted the proportions of students selecting the consistent alternatives in Form A and Form B to obtain the difference in the proportions selecting the consistent alternative from Form A to Form B. The average shift from Forms A to B was 0.31, again, a negligible shift.

Conclusions

The results of this study are consistent with those found by Huntley and Plake (1980): the cueing that is established by the generalizing of the stem is minimal. When cueing does appear to have an effect, it occurs within the vowel-consonant item types and functions, however subtly, differentially for males than females. This finding suggests that when working with the vowel-consonant item type, test constructors should

avoid the strategy of augmenting the item stem with vowel or consonant articles. Instead, rephrasing the items into neutral, yet grammatically correct, language would eliminate this potential source of cueing.

References

Board, C. & Whitney, D. R., The effect of selected poor item-writing practices on test difficulty, reliability and validity, Journal of Educational Measurement, 1972, 9, 3, 225-233.

Dudycha, A. L. & Carpenter, J. B., Effect of item format on item discrimination and difficulty, Journal of Applied Psychology, 1973, 58, 1, 116-121.

Dunn, T. F. & Goldstein, L. G., Test difficulty, validity, and reliability as functions of selected multiple-choice item construction principles, Educational and Psychological Measurement, 1959, 19, 2, 171-179.

Ebel, R. L., Essentials of Educational Measurement, Prentice-Hall, Englewood Cliff, New Jersey, 1972.

Hughes, H. H. & Trimble, W. E., The use of complex alternatives in multiple choice items, Educational and Psychological Measurement, 1965, 25, 1, 117-126.

Huntley, R. M. & Plake, B. S., Effect of selected item-writing practices on test performance: can relevant grammatical clues result in flawed items? Paper presented at the American Educational Research Association annual meeting, Boston, MA, 1980.

McMorris, R. F., Snyder, G. W., Brown, J. A., & Pruzek, R. M., Effect of violating item construction principles, Journal of Educational Measurement, 1972, 9, 4, 287-295.

Millman, J. & Setijadi. A comparison of the performance of American and Indonesian students on three types of test items. Journal of Educational Measurement, 1966, 59, 273-275.

Sarnacki, R. E. An Examination of Test-Wiseness in the Cognitive Test Domain. Review of Educational Research, 1979, 49, 2, 252-279.

Slatker, M. J., Koehler, R. A., & Hampton, S. H. Grade level, sex, and selected aspects of test wiseness. Journal of Educational Measurement, 1970, 1, 119-122.

Table 1

Distribution of Subjects by Sex

for Experimental Units One and Two

	Experimental Unit		
	One	Two	Total
Sex			
M	35	46	81
F	78	72	150
Total	113	118	231

Table 2

Item Score Means and Standard Deviation

by Form and Sex for Each Item Type

Plural-Singular

Form A (cue present)

		Item								
		1	2	3	4	5	6	7	8	9
Male	$\bar{X}$	2.66	1.85	2.37	2.59	2.00	2.76	2.66	2.72	2.00
	sd	1.211	0.51	1.00	1.04	1.11	1.08	0.87	1.05	1.21
Female	$\bar{X}$	2.77	2.10	1.71	2.65	2.10	2.72	2.73	3.07	2.73
	sd	0.88	0.95	1.06	1.08	1.06	0.94	1.09	1.04	1.14

Form B (cue absent)

Male	$\bar{X}$	2.76	1.80	1.98	2.65	2.04	2.77	2.70	2.91	2.78
	sd	0.85	0.63	1.06	1.08	1.05	0.77	0.78	1.01	1.35
Female	$\bar{X}$	2.74	2.13	1.99	2.63	1.94	2.63	2.50	2.82	2.97
	sd	.92	0.89	1.17	0.94	1.14	1.05	0.98	1.24	1.20

Table 2 (Cont.)

Vowel-Consonant

Form A (cue present)

		Item 1	2	3	4	5	6	7
Male	$\bar{X}$	2.80	2.14	3.04	2.51	3.33	2.91	3.57
	sd	1.24	0.55	0.92	0.92	0.92	1.01	0.86
Female	$\bar{X}$	2.82	2.21	3.10	2.33	3.25	2.15	3.17
	sd	1.16	0.63	1.02	1.06	0.82	1.09	1.11

Form B (cue absent)

Male	$\bar{X}$	2.71	2.15	3.20	2.20	3.31	2.63	3.29
	sd	1.15	0.59	0.87	0.78	1.16	0.97	0.99
Female	$\bar{X}$	2.95	2.76	3.08	2.51	3.17	2.35	3.08
	sd	1.27	0.65	0.88	0.99	1.11	1.23	1.07

Table 3

Two Factor (Form and Sex) ANOVA F Ratios

Plural-Singular and Vowel-Consonant Item Types

Source	Plural-Singular	Vowel-Consonant
Between		
Sex	.08	.53
Within		
Form	.26	1.26
Form X Sex	.88	0.02

Table 4

Cue Sensitivity for Total Group and By Sex

X^2 Goodness of Fit

Plural-Singular Item Type

		Alternatives				
		Inconsistent			Consistent	X^2 Value
		1	2	3	4	
	Expected	25.0	25.0	25.0	25.0	
	Total Group	25.2	23.1	31.3	20.0	2.73
Observed						
	Male	26.3	24.8	33.7	15.9	6.41
	Female	24.7	22.6	30.1	22.2	1.59

$_{.95}X^2_3 = 7.82$

Table 5

Cue Sensitivity for Total Group and By Sex

X^2 Goodness of Fit

Vowel-Consonant Item Type

		Alternatives				
		Inconsistent			Consistent	X^2 Value
		1	2	3	4	
	Expected	25.0	25.0	25.0	25.0	
	Total Group	16.0	26.6	38.3	18.6	11.6*
Observed	Male	11.1	28.8	40.7	18.8	19.7*
	Female	19.3	25.1	36.8	18.6	8.51*

*$_{.95}X^2 = 7.82$

Appendix 1

Examples of Item Types*

Singular-Plural Combination Type

Form A

1. Which foreign policy(ies) was(were) pursued by the United States after both world wars?

   I. Sending billions of dollars in foreign aid to Germany to assist in postwar reconstruction
   II. Furthering the interests of the U.S.S.R. and other communist nations
   III. Joining and actively supporting international peacekeeping organizations

   A. I only
   B. II only
   C. III only
   D. I, II, and III

2. Which situation(s) was(were) partly responsible for the outbreak of World War I?

   I. The economic rivalry between nations provoked by the Great Depression
   II. The expansionist policies of communist nations
   III. A long-standing territorial dispute between Japan and England over Hong Kong

   A. I only
   B. II only
   C. III only
   D. I, II, and III

* These items are examples of the content-free version of the experimental items; there is no correct answer.

Appendix 1, Continued

Singular-Plural Combination Type

Form B

1. After both world wars, United States foreign policy included:

   I. sending billions of dollars in foreign aid to Germany to assist in postwar reconstruction.
   II. furthering the interests of the U.S.S.R. and other communist nations.
   III. joining and actively supporting international peacekeeping organizations.

   A. I only
   B. II only
   C. III only
   D. I, II, and III

2. Which of the following situations contributed to the outbreak of World War I?

   I. The economic rivalry between nations provoked by the Great Depression
   II. The expansionist policies of communist nations
   III. A long-standing territorial dispute between Japan and England over Hong Kong

   A. I only
   B. II only
   C. III only
   D. I, II, and III

Appendix 1, Continued

Vowel-Consonant Type

Form A

1. The Free-Soil party was similar to the Republican party in that it advocated a(an):

   A. ban on immigration.
   B. purchase of disputed territory from Canada and Mexico.
   C. abolition of all taxes and tariffs.
   D. policy of guaranteeing adequate prices for farmers.

2. Any case that is tried by the Supreme Court must originate in a(an):

   A. act passed by Congress but vetoed by the president.
   B. conflict between the U.S. government and a private citizen.
   C. lawsuit involving at least one million dollars.
   D. conflict between federal and state laws or regulations.

Form B

1. The Free-Soil party was similar to the Republican party in that it advocated a:

   A. ban on immigration.
   B. purchase of disputed territory from Canada and Mexico.
   C. prohibition of all taxes and tariffs.
   D. policy of guaranteeing adequate prices for farmers.

2. Any case that is tried by the Supreme Court must originate in:

   A. bills passed by Congress but vetoed by the president.
   B. conflicts between the U.S. government and a private citizen.
   C. lawsuits involving at least one million dollars.
   D. conflicts between federal and state laws or regulations.